

UNITED STATES DISTRICT COURT  
FOR THE DISTRICT OF MASSACHUSETTS

SINGULAR COMPUTING LLC,  
Plaintiff,  
v.  
GOOGLE LLC,  
Defendant.

Civil Action No. 1:19-cv-12551 FDS  
Hon. F. Dennis Saylor IV

**DEFENDANT GOOGLE LLC'S OPPOSITION TO SINGULAR'S**  
**MOTION IN LIMINE NO. 7**

## I. INTRODUCTION

This motion is an attempt to prevent the jury from learning about significant errors that Singular’s damages expert, Mr. Green, made in calculating his reasonable royalty in this case. Among other errors, Mr. Green conducts a flawed apportionment analysis, which fails to ensure that his “ultimate reasonable royalty award [is]... based on the incremental value that the patented invention adds to the end product.” *Commonwealth Sci. & Indus. Rsch. Org. v. Cisco Sys., Inc.*, 809 F.3d 1295, 1301 (Fed. Cir. 2015). Mr. Green also errs in his decision to use GPUs using the *fp32* number format as a basis of his cost-savings damages calculations, instead of using GPUs using the *fp16* number format. Such errors artificially and significantly inflate Mr. Green’s royalty numbers.

As part of her opinions explaining the errors in Mr. Green’s damages opinion, Google’s damages expert—Ms. Stamm—conducts an alternative apportionment analysis under the analytical approach. To conduct her alternative analysis, Ms. Stamm relies on a benchmark comparison between the use of *bf16* and the use of *fp16* (two 16-bit number formats) — an analysis that Ms. Stamm explains more effectively isolates and determines the value of the Asserted Claims than Mr. Green’s erroneous opinion based on the use of *fp32*.<sup>1</sup> Singular never filed a *Daubert* motion on this apportionment analysis in Ms. Stamm’s damages opinion. Despite this, Singular seeks to use this motion *in limine* to preclude Ms. Stamm from presenting her explanation of why Singular’s damages theory is fatally flawed. Singular is also seeking more broadly to limit Google’s ability to cross examine Singular’s damages expert, Mr. Green, on his flawed damages and apportionment theories and to limit Google’s ability to submit evidence showing why Mr. Green’s theories are flawed.

Singular’s motion *in limine* No. 7 should be denied for the following reasons. **First**, Singular’s motion regarding Ms. Stamm’s opinion is a thinly-disguised and untimely *Daubert*

---

<sup>1</sup> Ms. Stamm’s only non-infringing alternative calculation (premised on TPUs implementing *bf20*) is not at issue in this motion.

motion. **Second**, Singular cannot exclude Ms. Stamm’s *fp16* analysis on the grounds that it is an improper “non-infringing alternative” because it is not. As Ms. Stamm has repeatedly explained, her *fp16* analysis is an alternative apportionment analysis permitted by the analytical approach that shows the errors in Mr. Green’s apportionment analysis. **Third**, Singular presents no legitimate basis to preclude Google from presenting evidence or argument regarding *fp16* in this case – a preclusion that would improperly limit highly relevant cross-examination and argument and deprive Google its right to rebut Mr. Green’s flawed assumptions and damages calculations. Singular’s motion should be denied.

## II. FACTUAL BACKGROUND

**Google’s TPU Systems and Chips:** Google’s Tensor Processing Unit (“TPU”) systems are designed for machine learning computing tasks. Two versions of Google’s TPU systems are accused here: TPUv2 and TPUv3. These TPU systems consist of multiple TPU chips. Each TPU chip and system undisputedly contains many components unrelated to the asserted claims. Although Singular’s infringement allegations target functionality specific to *subparts* of the TPU v2 and v3 *chips*, Mr. Green’s royalty base is improperly derived from figures relating to the broader TPU systems. *See* Dkt. 477 at 2-3.

**Singular’s Damages Theory:** Singular’s damages calculations are based on a reasonable royalty analysis supposedly informed by both “analytical approaches and the results of a hypothetical negotiation.” Ex. 1 (Green Rpt.) at 4. A fundamental premise of Mr. Green’s damages theory is that, if Google were unable to use Singular’s technology, Google would have incurred increased “costs” by entirely replacing its TPU systems with third-party graphics processing unit (“GPU”) systems. Ex. 1 (Green Rpt.) at 63, Section V.C. To calculate these allegedly saved costs, Mr. Green assumes that instead of using TPUs, Google would have relied on GPUs using a 32-bit number format called *fp32* rather than a 16-bit floating-point number format called *fp16*. Mr. Green ignores the *fp16* number format even though Singular’s technical expert expressed the opinion that *neither fp32 nor fp16* would infringe the asserted claims.

Mr. Green also errs by conducting a flawed, “top down” apportionment, in which he starts with a royalty base premised on entire TPU systems. *See, e.g.*, Ex. 1 (Green Rpt.) at 107; *see also* Dkt. 477. These errors (and others) lead Mr. Green to posit flawed royalty numbers worth [REDACTED]. *Id.* at 105.

**Google’s Damages Rebuttal:** In rebuttal to Mr. Green’s damages opinion, Google’s damages expert, Ms. Stamm, proffers one alternative damages calculation based upon a non-infringing alternative – the use of a number format called *bf20*. Ex. 2 (Stamm Rpt.) at 4. Ms. Stamm also conducts a different, and separate, “bottom up” apportionment analysis as one method of showing the errors in Mr. Green’s flawed apportionment approach and damages framework.<sup>2</sup> As Ms. Stamm states in her report, Singular’s 40% apportionment approach is wrong because it mistakenly uses *fp32* as a benchmark instead of *fp16* for GPUs, *id.* at 61, thus artificially and significantly inflating the supposed value of the Asserted Claims by choosing a slower alternative.

To counter Mr. Green’s flawed apportionment theory, Ms. Stamm provides an alternative analysis that excludes the performance improvement between the use of GPU chips with the *fp32* number format as compared to the *fp16* number format. To more directly and accurately measure the incremental value contributed by the asserted claims (as opposed to unrelated contributions), Ms. Stamm utilizes this alternative apportionment method as “a better means for isolating the contribution of the Asserted Claims” by instead “comparing the use of *bf16* . . . with the use of *fp16*” to evaluate “the differences attributable only to the different number formats used.”<sup>3</sup> *Id.* at 67. Specifically, Ms. Stamm calculates the costs Google would incur to achieve

---

<sup>2</sup> As explained in Ms. Stamm’s report, Mr. Green’s apportionment analyses “are flawed and overstate the contribution of the Asserted Claims,” because Mr. Green includes a “myriad of [unrelated] benefits to Google from using TPU systems.” *See id.* at 60, 70; *see also id.* Section IX.I.

<sup>3</sup> Both parties agree that these different number formats – *fp16* and *bf16* – both have 16 bits and similar performance when used for matrix multiplication for AI; however, the use of *fp16* would not infringe the Asserted Claims.

the same performance using the *fp16* number format to measure the incremental value of using the *bf16* number format for matrix multiplication.

Ms. Stamm's report is clear that her use of the *fp16* benchmark is not provided as "the next best non-infringing alternative," but is instead provided as an alternative apportionment method designed to "effectively determine[] the value of the Asserted Claims by capturing only those cost savings related to the use of *bf16* as compared to *fp16*." *Id.* at 70. Ms. Stamm uses this alternative apportionment analysis to show how Mr. Green's apportionment analysis is significantly inflated.

**Singular's *Daubert* Motion:** The Court set a deadline of April 28, 2023 for *Daubert* motions. Dkt. No. 458. Under that deadline, Singular filed a *Daubert* motion attacking one of Ms. Stamm's opinions, but not the opinion at issue in this motion *in limine*. *See* Dkt. No. 466-1. In that *Daubert* motion, Singular moved to exclude Ms. Stamm's non-infringing alternative damages opinion, an opinion premised on TPUs using the *bf20* number format. *Id.* Singular's motion did not seek to exclude Ms. Stamm's alternative apportionment analysis based on TPUs using the *fp16* number format. *Id.* Google opposed that motion (Dkt. 509), which is still pending.

### III. ARGUMENT

#### A. Singular's motion is untimely and improper at this stage.

Singular's motion primarily seeks to exclude an opinion, and limit the testimony, of Google's damages expert Ms. Stamm. Singular's motion should be denied because it is untimely.

The Court set a deadline of April 28, 2023 for *Daubert* motions. Dkt. No. 458. Singular filed a *Daubert* motion as to Ms. Stamm, seeking to exclude her opinion regarding the bfloat20 non-infringing alternative. *See* Dkt. No. 466-1. Significantly, Singular's *Daubert* motion did *not* take issue with the portion of Ms. Stamm's opinion at issue in this motion regarding

apportionment and *fp16*, *see* Dkt. No. 466-1, even though all of the information Singular seeks to exclude now was known to Singular at the *Daubert* deadline.

Singular should not be permitted a second bite at the *Daubert* apple through a thinly disguised motion *in limine* because any such motion is untimely. *See, e.g., Great Am. Ins. Co. of N.Y. v. Midland Chin Baptist Church*, No. MO:20-CV-25, 2022 WL 18046718, at \*2 (W.D. Tex. Feb. 3, 2022) (denying motion *in limine* regarding expert testimony in part because “the deadline for *Daubert* motions ha[d] passed”); *Rybas v. Riverview Hotel Corp.*, No. ELH-12-03103, 2015 WL 11027598, at \*3 (D. Md. Feb. 4, 2015) (denying motion *in limine* seeking to exclude expert testimony as an untimely *Daubert* challenge requiring “a preliminary assessment by the Court as to the methodology, foundation, and scientific validity of the expert’s opinions”); *Penthouse Owners Assoc., Inc. v. Certain Underwriters at Lloyd’s, London*, No. 1:07-CV-568-HSO-RSW, 2011 WL 13073684, at \*2 (S.D. Miss. Feb. 9, 2011) (denying motion *in limine* to exclude expert testimony and finding challenge to expert testimony waived after deadline for *Daubert* motions had passed). In fact, much of the inapposite case law Singular cites in this motion regarding acceptability and availability is the *same* case law Singular cited in its *Daubert* motion regarding Google’s *bf20* non-infringing alternative. This underscores that Singular could have filed a *Daubert* motion on the issues presented in this motion, but chose not to.

Thus, the Court should deny this motion because it amounts to a belated *Daubert* motion “without the protections and requirements of such a motion.” *SPEX Techs. v. Apricorn, Inc.*, No. CV 16-07349-JVS (AGRx), 2020 WL 1289546, at \*1 (C.D. Cal. Jan. 21, 2020) (denying motion *in limine* that was a disguised motion for summary judgment). Singular’s motion also does not provide any non-*Daubert* basis – evidentiary or otherwise – to exclude testimony or argument about Ms. Stamm’s apportionment analysis.<sup>4</sup>

---

<sup>4</sup> Singular asserts in its motion that “no reasonable jury could conclude that *fp16* was available at the time of the hypothetical negotiation.” *See, e.g.*, Mot. at 9. Although Singular invokes this standard – which is typically applied to motions for summary judgment, Singular provides no basis for why the Court should apply this standard for a motion *in limine*.

**B. Ms. Stamm’s alternative apportionment analysis is appropriate and highly relevant to an accurate reasonable royalty determination.**

Singular’s motion *in limine* seeks to exclude Ms. Stamm’s *fp16*-based apportionment analysis because Singular contends the analysis constitutes an improper, “non-infringing alternative.” *See, e.g.*, Mot. at 1, 4. Singular is wrong, and its motion should be denied for this reason as well.

At trial, Singular’s damages expert apparently intends to testify that, if Google had been unable to use the allegedly infringing functionality in the TPU chips, Google would have stopped using TPU systems entirely and instead would have used third party GPU systems. *See Ex. 1* (Green Rpt.) at Section V.C. Mr. Green’s cost-savings calculations are premised on this theory, including his erroneous follow-on assumption that, if Google had replaced TPUs with GPUs, it would have used GPUs implementing a 32-bit number format called *fp32* instead of a 16-bit number format, like *fp16*.

To counter Mr. Green’s apportionment opinions (and demonstrate his errors), Google’s expert – Ms. Laura Stamm – applies the analytical approach. Unlike Mr. Green’s impermissible “top-down” apportionment approach that fails to properly isolate the incremental value of these patents, *see Dkt. 477*, Ms. Stamm isolates the contribution of the Asserted Claims by conducting an apportionment analysis that starts from the “bottom up.” Thus, Ms. Stamm evaluates “the differences attributable only to the different number formats used for matrix multiplication” — *bf16* and *fp16*.<sup>5</sup> Ex. 2 (Stamm Rpt.) at 67. Ms. Stamm’s alternative apportionment analysis more directly and accurately measures the incremental value of the asserted claims by excluding the unrelated benefits Mr. Green subsumes by comparing GPU chips using *fp32* instead of *fp16*. This alternative calculation is an appropriate method to isolate the incremental value of the

---

<sup>5</sup> As Singular’s technical expert Dr. Khatri concedes, these different number formats – *fp16* and *bf16* – both have 16 bits and similar (but not identical) performance when used for matrix multiplication for machine learning; however, he opines that the use of *fp16* would not infringe the Asserted Claims. *See Ex. 3* (Khatri Rpt.) at 22 (chart asserting that *fp16* does not satisfy the claim language).

Asserted Claims, thus satisfying the apportionment requirement. *See VirnetX Inc. v. Cisco Sys.*, 767 F.3d 1308, 1326 (Fed. Cir. 2014) (“[A] patentee must take care to seek only those damages attributable to the infringing features.”).

Ms. Stamm’s use of the analytical approach is a proper and appropriate apportionment methodology, especially in a case involving a multi-component product.<sup>6</sup> *See* Mot. at 2 (recognizing that Ms. Stamm’s *fp16* apportionment calculations are “consistent with the well accepted cost-savings methodology”). As courts recognize when a valid *Daubert* motion is filed to present the question of whether a reasonable royalty amount damages opinion can be precluded pre-trial (not the situation here), “[t]he essential requirement is that the ultimate reasonable royalty award must be based on the incremental value that the patented invention adds to the end product.” *Ericsson, Inc. v. D-Link Sys. Inc.*, 773 F.3d 1201, 1226 (Fed. Cir. 2014). In light of this, courts routinely allow experts to use different, reliable methods for calculating a royalty, including analytical approaches using similar benchmark comparisons. *See, e.g., Metaswitch Networks Ltd. v. Genband US LLC*, No. 2:14-cv-744-JG-RSP, 2016 WL 874737, at \*4–5 (E.D. Tex. Mar. 5, 2016) (declining to exclude analytical approach so long as the comparison isolates the value of the patented features—and no more—in comparing two products as benchmarks). Ms. Stamm’s methodology for an alternative apportionment is an economically and logically sound method to apportion the incremental value of the Asserted Claims without subsuming additional, unrelated benefits. And it is an appropriate opinion to provide to the jury as a counter to show why Mr. Green’s apportionment analysis is flawed.

Singular’s motion ignores the merits of Ms. Stamm’s apportionment analysis and instead attempts to recast it as a “non-infringing alternative” opinion. *See, e.g., Mot.* at 1, 4. This is

---

<sup>6</sup> Mr. Green concedes that the analytical approach of benchmarking is appropriate in his report. Ex. 1 (Green Rpt.) at 3 (“‘analytical approaches’ can be used to determine a royalty rate...[t]ypically, these methods compare the benefit or returns the infringer has obtained from the use of the patented technology with its other products, or some other benchmark. . . .”); *see also id.* at 60 (describing the analytical approach as a method comparing the patented technology to other benchmarks).

wrong – Ms. Stamm has been clear both in her report and at her deposition that the *fp16* apportionment analysis she provides is her “alternative method of isolating [the] value of the asserted claims” which is being proffered to explain why Mr. Green’s apportionment analysis is wrong:

[REDACTED]

Ex. 2 (Stamm Rebuttal Rpt.) at 70; *see also* Ex. 4 (Stamm Depo. Tr.) at 57:15-24 [REDACTED]

[REDACTED]. Ms. Stamm was also clear in both her report and deposition that her *fp16* apportionment analysis is not being proffered as a non-infringing alternative, including when she states that:

[REDACTED]

*Id.*; Ex. 4 (Stamm Depo. Tr.) at 61:25-62:5 [REDACTED]

[REDACTED] *see also id.* at 97:4-10.<sup>7</sup> Given these clear statements to the contrary, all of Singular’s cited case law regarding the acceptability or availability of a non-infringing alternative is inapposite; Ms. Stamm’s *fp16* analysis apportions

---

<sup>7</sup> Singular’s reference to the deposition of Google’s technical expert – Dr. Martin Walker – is a red herring. Dr. Walker described *fp16* as a non-infringing alternative in the technical sense of the term. The use of *fp16* for matrix multiplication would not infringe the Asserted Claims and would serve as an alternative in terms of available options; however, Dr. Walker does not determine whether *fp16* is the next best non-infringing alternative or used to conduct an apportionment analysis for damages purposes. That falls in Ms. Stamm’s purview.

and isolates the value of the Asserted Claims using the analytical approach and is not being used as a non-infringing alternative.

**C. Singular cannot exclude questioning or argument regarding *fp16*.**

Singular also cannot use this motion *in limine* to limit Google's evidence and argument regarding *fp16* or to limit Google's cross-examination of Mr. Green on his erroneous assumptions regarding *fp32* and his failures to consider *fp16*. Singular's request to exclude *any* evidence or argument regarding *fp16* should therefore be denied.<sup>8</sup>

The sole basis of Singular's broad request to exclude "all evidence concerning *fp16*" is Singular's claim that "no reasonable jury could conclude that *fp16* was available at the time of the hypothetical negotiation." *See* Mot. at 9-12. However, the only case law that Singular cites to support this claim is case law addressing whether evidence meets the standard for "availability" applicable to a non-infringing alternative. *See id.* The question of whether *fp16* was sufficiently available in 2017<sup>9</sup> to satisfy the non-infringing alternative standard does not provide an evidentiary basis to broadly exclude all evidence regarding *fp16* in this case.

**First**, as explained above, Mr. Green mistakenly selects V100 GPUs using *fp32* as the benchmark for both his cost-savings analysis and his apportionment analysis. Google and its experts contend that even under Mr. Green's incorrect approach, the appropriate benchmark should at least have been V100 GPUs using *fp16*. Because this assumption impacts the rest of Mr. Green's analysis and significantly inflates his royalty opinions, Google has a right to rebut this issue, to attack the factual assumptions underlying Mr. Green's analysis, and to explain to

---

<sup>8</sup> Singular's motion expressly asks the Court to prevent Google from presenting "evidence or argument concerning the costs associated with designing, developing, deploying or using TPUs or GPUs running *fp16* instead of *bf16*." *See* Mot. at 1-2 (intro), 12 (conclusion). However, in its argument Section B Singular buries a far broader request, asking the Court to exclude "all evidence concerning *fp16*." *Id.* at 9-12. If the Court grants this motion, Singular should be limited to the narrower scope of relief sought in both its introduction and conclusion.

<sup>9</sup> There's no dispute that *fp16* as a number format was available in 2017 and for many years before 2017; as Mr. Green concedes in his report, the Institute of Electrical and Electronics Engineers approved standard 754-2008 in 2008, establishing *fp16* as a format used in the industry to represent numbers for computing. *See* Ex. 1 (Green Rpt.) at 14-16.

the jury why those assumptions were not reasonable. The jury should have the opportunity to understand that Mr. Green's incorrect assumptions have massive implications on his analysis.

**Second**, even if there were a basis to apply case law directed at a non-infringing alternative to the broader question of whether evidence regarding *fp16* can be presented at trial (there is not), the question of whether an alternative design was “available” at the time of the hypothetical negotiation is one of weight, not admissibility, and should be decided by the jury. *See Verinata Health, Inc. v. Ariosa Diagnostics, Inc.*, No. 12-CV-05501-SI, 2018 WL 288050, at \*5 (N.D. Cal. Jan. 4, 2018) (finding that objections to the availability of non-infringing alternatives “go to the weight rather than the admissibility of the evidence” and declining to preclude such testimony); *Greatbatch Ltd. v. AVX Corp.*, No. CV 13-723-LPS, 2015 WL 9171042, at \*7 (D. Del. Dec. 11, 2015) (same), *aff'd*, 813 F. App'x 609 (Fed. Cir. 2020); *Carnegie Mellon Univ. v. Marvell Tech. Grp., Ltd.*, Civ. No. 09-290, 2012 WL 3686736, \*4–5 (W.D. Pa. Aug. 24, 2012) (denying motion to exclude two damages experts over patentee's objections that the experts did not know if the purported non-infringing alternatives had actually been built, tested or sold, and at what costs, as those details went to the factual question of the alternatives' existence, not the admissibility of the opinions). Here, as explained in Google's opposition to Singular's motion *in limine* No. 9, ample evidence exists to show that *fp16* did in fact exist in March 2017 and was being used for training machine-learning models. Google's Opposition to Singular's MIL No. 9 at 2, 3-4. Indeed, Mr. Green's damages report relies on numerous articles that discuss the use of *fp16*, both on GPUs and for machine-learning applications (though Mr. Green chose to ignore those portions of the articles). *Id.* Singular's attempts to exclude this evidence highlights why this issue of fact – underlying Mr. Green's significantly flawed assumptions – should be decided by the jury.

The jury is entitled to reach its own conclusion regarding acceptability, availability, and credibility based on all of the evidence. Singular's motion to prevent Google from discussing or highlighting Mr. Green's flawed opinion should be denied.

Respectfully submitted,

Dated: December 13, 2023

By: /s/ Nathan R. Speed  
Gregory F. Corbett (BBO #646394)  
gcorbett@wolfgreenfield.com  
Nathan R. Speed (BBO #670249)  
nspeed@wolfgreenfield.com  
Elizabeth A. DiMarco (BBO #681921)  
edimarco@wolfgreenfield.com  
Anant K. Saraswat (BBO #676048)  
asaraswat@wolfgreenfield.com  
WOLF, GREENFIELD & SACKS, P.C.  
600 Atlantic Avenue  
Boston, MA 02210  
Telephone: (617) 646-8000  
Fax: (617) 646-8646

Robert Van Nest (admitted *pro hac vice*)  
rvannest@keker.com  
Michelle Ybarra (admitted *pro hac vice*)  
mybarra@keker.com  
Andrew Bruns (admitted *pro hac vice*)  
abrunz@keker.com  
Vishesh Narayen (admitted *pro hac vice*)  
vnarayen@keker.com  
Christopher S. Sun (admitted *pro hac vice*)  
csun@keker.com  
Anna Porto (admitted *pro hac vice*)  
aporto@keker.com  
Deeva Shah (admitted *pro hac vice*)  
dshah@keker.com  
Stephanie J. Goldberg (admitted *pro hac vice*)  
sgoldberg@keker.com  
Eugene M. Paige (admitted *pro hac vice*)  
epaige@keker.com  
Rachael E. Meny (admitted *pro hac vice*)  
rmeny@keker.com  
Kaiyi A. Xie (admitted *pro hac vice*)  
kxie@keker.com  
Eric K. Phung (admitted *pro hac vice*)  
ephung@keker.com  
Spencer McManus (admitted *pro hac vice*)  
smcmanus@keker.com  
KEKER, VAN NEST & PETERS LLP

633 Battery Street  
San Francisco, CA 94111-1809  
(415) 391-5400

Michael S. Kwun (*pro hac vice*)  
mkwun@kblfirm.com  
Asim Bhansali (*pro hac vice*)  
abhangsali@kblfirm.com  
KWUN BHANSALI LAZARUS LLP  
555 Montgomery Street, Suite 750  
San Francisco, CA 94111  
(415) 630-2350

Matthias A. Kamber (*pro hac vice*)  
matthiaskamber@paulhastings.com  
PAUL HASTINGS, LLP  
101 California Street  
Forty-Eighth Floor  
San Francisco, CA 94111  
(415) 856-7000

Ginger D. Anders (*pro hac vice*)  
Ginger.Anders@mto.com  
J. Kain Day (*pro hac vice*)  
kain.day@mto.com  
MUNGER, TOLLES & OLSON LLP  
601 Massachusetts Avenue NW, Suite 500E  
Washington, D.C. 20001  
(202) 220-1100

Jordan D. Segall (*pro hac vice*)  
jordan.segall@mto.com  
MUNGER, TOLLES & OLSON LLP  
350 South Grand Avenue, 50th Floor  
Los Angeles, CA 90071-3426  
(213) 683-9100

*Counsel for Defendant Google LLC*

**CERTIFICATE OF SERVICE**

I certify that this document is being filed through the Court's electronic filing system, which serves counsel for other parties who are registered participants as identified on the Notice of Electronic Filing (NEF). Any counsel for other parties who are not registered participants are being served by first class mail on the date of electronic filing.

*/s/ Nathan R. Speed* \_\_\_\_\_

Nathan R. Speed